BAB I
PENDAHULUAN

1.1 Latar Belakang

Seiring dengan bertambahnya kapasitas dan trafik jaringan internet dudunia, serangan terhadap keamanan jaringan komputer menjadi meningkat (Corporation, 2015). Peningkatan serangan jaringan tersebut menyebabkan data yang harus dianalisis menjadi sangat banyak, sehingga sistem keamanan yang ada mengalami keterbatasan dalam beradaptasi dengan jumlah data yang besar dan serangan baru. Untuk mengatasi permasalah tersebut, maka dibutuhkanlah sebuah sistem keamanan yang dapat membantu menganalisis data yang besar dan dapat menemukan serangan baru (Bloedorn dkk, 2001).

IDS merupakan sebuah kemampuan yang dimiliki oleh sebuah sistem atau perangkat untuk dapat melakukan deteksi terhadap serangan yang mungkin terjadi dalam jaringan komputer (Izza, Khaerani. Lekso, 2015). Secara umum terdapat beberapa cara bagaimana IDS bekerja, yaitu IDS berbasis *signature* dan IDS berbasis *anomaly*. IDS berbasis *signature* bekerja dengan melakukan pencocokan lalu lintas jaringan dengan pola serangan yang ada (seperti halnya antivirus), kekurangan dari IDS ini yaitu tidak dapat mendeteksi serangan yang tidak diketahui dan perlu memperbarui tanda pola serangan ketika ada serangan baru. IDS berbesis *Anomaly* bekerja dengan cara menangkap semua informasi *header* dari paket IP lalu mengidentifikasi apakah informasi tersebut menyimpang dari perilaku normal (Seth ,2017). IDS berbasis *anomaly* akan mengklasifikasikan serangan setelah mempelajari pola serangan, oleh karena itu kinerja metode klasifikasi merupakan faktor penting yang mempengaruhi kinerja IDS. (Garg dan Khurana, 2014).

Berdasarkan penelitian yang dilakukan oleh Tavallaee dkk pada tahun 2010, yang menganalisis beberapa penelitian tentang IDS dari tahun 2000-2008, mengatakan bahwa hampir 70% penelitian dilakukan menggunakan kumpulan data NSL-KDD

(*Network Security Layer - Knowledge Discovery in Database*), dan secara khusus berhasil mencapai tujuan penelitinya. Mahbod dkk pada tahun 2009, juga mengatakan Dataset NSL-KDD merupakan kumpulan data benchmark yang efektif membantu peneliti membandingkan metode-metode deteksi instrusi sebagai langkah awal membangun *Intrusion detection system (IDS)*.

Walaupun dataset NSL-KDD banyak digunakan, namun masih terdapat permasalahan pada Dataset tersebut. Mukherjee dan Sharma tahun 2012, mengatakan bahwa dataset NSL-KDD memilki fitur yang tidak relevan dan berlebihan, sehingga akan mengakibatkan proses pendeteksian yang panjang dan menurunkan kinerja sistem deteksi intrusi (IDS). Ambusaidi dkk tahun 2016 juga mengatakan bahwa, dataset NSL-KDD memiliki fitur noisy, berlebihan atau tidak informatif, mengakibatkan algoritma klasifikasi yang ada mengalami *Detection rate* dan *False positif rate* yang rendah.

Untuk mengatasi permasalah yang terdapat pada dataset NSL-KDD tersebut, langkah yang dilakukan oleh peneliti-peneliti lain adalah dengan menerapakan *feature selection*. *Feature selection* adalah proses pemilihan atribut yang sesuai dari kumpulan atribut pada dataset NSL, seperti yang dilakukan oleh Mukherjee dan Sharma tahun 2012, Deshmukh dkk tahun 2014, Garg dan Kumar tahun 2014, dan Deshmukh dkk tahun 2015. Hasil dari penelitian diatas mengatakan bahwa penerapan *feature selection* berhasil meningkatkan akurasi dari metode yang mereka gunakan.

Mukherjee dan Sharma tahun 2012, melakukan penelitian dengan menerapkan *feature selection CfsSubsetEval, Gain Ratio, Information Gain* dan *Feature Vitality Based Method* pada dataset NSL-KDD menggunakan metode klasifikasi Naïve Bayes. Hasil penelitiannya menyatakan bahwa penerapan *feature selection* berhasil meningkatkan hasil akurasi dari metode Naïve Bayes.

Deshmukh dkk pada tahun 2014, melakukan penelitian dengan penerakan *feature selection CfsSubsetEval* pada dataset NSL-KDD menggunakan metode klasifikasi Naïve Bayes, NB Tree, dan AD Tree. Dari hasil penelitian, diketahui bahwa penerapan *feature selection* berhasil meningkatkan hasil akurasi setiap metode

klasifikasi, namun metode NB Tree dan AD Tree membutuhkan waktu yang lama untuk proses klasifikasi dibandingkan metode Naïve Bayes. Naïve Bayes memiliki keunggulan yaitu lebih cepat dan efisien ruang dan lebih sederhana dibandingkan metode NB Tree dan AD Tree.

Berhubungan dengan penerapan *feature selection*, Garg dan Kumar pada tahun 2014 melakukan penelitian dengan mengkombinasikan *feature selection* menggunakan Dataset NSL-KDD dengan jumlah data latih yaitu 93900 dan 48372 data uji yang diambil secara acak. *Feature selection* yang digunakan adalah *Information Gain, Gain Ratio, ReliefF, OneR, Symmetrical Uncertainty, Chi-Square*, kemudian dievaluasi mengunakan 10 metode klasifikasi yaitu *Rotation Forest, Random Tree, Random Committee, Random Forest, IBK, Random Sub Space, IB1, Part, Jrip, NB Tree*. Hasil penelitian ini menyatakan bahwa kombinasi *Symmetrical Uncertainty* dan *Gain Ratio* memeberikan hasil akurasi yang baik pada metode IBK dengan akurasi 98.5%.

Berdasarkan hasil dari beberapa penelitian terkait tersebut, pada penelitian kali ini penulis mengusulkan penerapan *feature selection Symmetrical Uncertainty* dan *Gain Ratio* pada dataset NSL-KDD menggunakan metode klasifikasi *Naïve Bayes*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan diatas, rumusan masalah pada penelitian ini adalah bagaimana mengimplementasikan kombinasi hasil *feature selection Gain Ratio* dan *Symmetrical Uncertainty* pada dataset NSL-KDD menggunakan metode klasifikasi *Naïve Bayes*.

1.3 Batasan Masalah

Dalam melakukan sebuah penelitian, diperlukan batasan-batasan agar tidak menyimpang dari perencanaan. Adapun batasan masalah pada penelitian ini adalah :

1. Data yang akan digunakan pada penelitian ini adalah dataset NSL-KDD latih dengan jumlah data 208 yang diambil secara acak dan merata sesuai jumlah data terkecil pada kelas dataset NSL-KDD.
2. Jumlah atribut dataset NSL-KDD yang akan digunakan adalah 41 atribut dan Kelas yang digunakan adalah DoS, R2L, Probe dan U2R. Dengan jumlah data pada setiap kelasnya adalah 52
3. *Feature selection* yang akan digunakan adalah *Gain Ratio* dan *Symmetrical Uncertainty* yang akan analisa menggunakan *tools* Weka 3.8.
4. Pada tahap *pre-processing* dataset NSL-KDD akan dilakukan menggunakan teknik *Discretization* 10 interval yang dianalisa menggunakan *tools* Weka 3.8.

1.4 Tujuan Penelitian

Tujuan penelitian yang ingin dicapai penulis adalah menerapkan kombinasi hasil *Feature Selection* yaitu *Gain Ratio* dan *Symmetrical Uncertainty* pada dataset NSL-KDD menggunakan metode klasifikasi *Naïve Bayes*.

1.5 Sistematika Penulisan

Sistematika penulisan Tugas Akhir ini memberikan gambaran umum tentang penelitian yang dilakukan. Sistematika penulisan ini dibagi menjadi beberapa bab, yaitu :

BAB I PENDAHULUAN

Bab ini membahas tentang gambaran umum isi tugas akhir yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menguraikan tentang teori-teori berhubungan yang digunakan untuk menganalisa masalah dan teori yang digunakan dalam mengolah data penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini membahas tentang langkah-langkah yang dilakukan dalam proses penelitian yaitu dimulai dari persiapan penelitian, pengumpulan data, analisa, perancangan, implementasi hingga pengujian.

BAB IV ANALISA DAN PERANCANGAN

Bab ini berisi pembahasan mengenai analisa data sesuai dengan tahapan *Knowledge Discovery in Databases* (KDD) dan akan membahas perancangan terhadap sistem yang akan dibuat sesuai dengan model klasifikasi yang digunakan.

BAB V IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang lingkungan implementasi dan hasil inplementasi dari perancangan yang telah dilakukan, serta pengujian terhadap sistem menggunakan *Blackbox testing* dan pengujian terhadap akurasi metode klasifikasi *Niave Bayes* menggunakan tabel *Confusion Matrix*.

BAB VI KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang dihasilkan dari pembahasan implementasi dan pengujian pada bab sebelumnya dan saran untuk penelitian dimasa yang akan datang.